DOCUMENT RESUME

ED 370 998 TM 021 607

AUTHOR Croft, Cedric
TITLE The Conflicting World of Standards-Based Assessment.
PUB DATE Dec 93
NOTE 20p.; Paper presented at the National Conference of the New Zealand Association for Research in Education (15th, Hamilton, New Zealand, December 2-5, 1993).
AVAILABLE FROM Information Service, New Zealand Council for Educational Research, P.O. Box 3237, Wellington, New Zealand.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Competence; *Criterion Referenced Tests; *Educational Assessment; Elementary Secondary Education; Evaluation Methods; Foreign Countries; Measurement Techniques; *National Competency Tests; Needs Assessment; Norm Referenced Tests; Reliability; Research Needs; Skill Development; Standards; Validity
IDENTIFIERS National Qualifications Framework (New Zealand); *New Zealand

ABSTRACT

Standards-based assessment, or at least the concept of standards-based assessment, will provide a key strategy for the implementation of the New Zealand National Qualifications Framework. This paper considers the meaning of standards-based assessment and its role in New Zealand's assessment for nationally recognized qualifications. Standards-based assessment, which can be divided into competency-based and achievement-based assessments, is distinguished from norm-referenced assessment in that the measurement or outcome is assessed against some fixed criterion or level of achievement known as a standard. Competency-based assessment is then based on the presence or absence of a set of skills, while achievement-based assessment refers to the extent to which skills are present. Regardless of whether New Zealand has overemphasized the importance of standards-based assessment over norm-referenced assessment, establishing the validity and reliability of standards-based assessment and researching the questions that will enhance these qualities would seem to be a priority. (Contains 12 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Available from: Information Service,
New Zealand Council for Educational
Research, P.O. Box 3237, Wellington,
New Zealand (Fax 64 4 3847933)

TM

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

K.A. Pickens

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED 370 998

The Conflicting World of Standards-Based Assessment

Cedric Croft
NZCER

Paper Delivered at the 15th National Conference of
New Zealand Association For Research in Education,
Hamilton, 2-5 December 1993

TM 021 607

BEST COPY AVAILABLE

Introduction

Although a relative newcomer to the assessment scene in New Zealand, standards-based assessment, or at least the concept of standards-based assessment, will provide a key strategy for the implementation of the National Qualifications Framework. An Introduction to the Framework, p. 6 states,

> "In the Framework, assessment for all new nationally recognised qualifications will be based on standards which will have been agreed on and set by industry for vocational qualifications, and by appropriate professional groups for general education. This means that a learner's performance will be measured against clearly stated and well defined standards of achievement or competence."

Clearly, standards will provide a frame of reference for assessing learning outcomes contained in a multiplicity of Unit Standards. Additionally, the 'range of assessments' included in any Unit Delivery, (p. 5, Introduction) will be dictated by the theoretical and practical possibilities encompassed by standards-based assessment. Hence, standards-based assessment is to provide a key strategy if not the key strategy, for the successful implementation of the Framework. Given this central position, it may be seen as surprising that standards-based assessment has not been the subject of more clarification, debate or discussion.

What is Standards-Based Assessment?

A logical place to begin is to consider the simple question 'What is standards-based assessment'?

In an NZQA publication, Beyond The Norm? An Introduction to Standards-based Assessment (Peddie, 1992) it is stated p. 21,

> "Material published by the Qualifications Authority to date draws a clear distinction between two main types of

assessment, norm-referenced and standards-based. Standards-based assessment is then divided into competency-based and achievement-based assessment."

The distinction referred to between norm-referenced and standards-based assessment is important, and worth looking at further.

At the broadest level we can distinguish between Comparative and Non-Comparative assessment, Withers and Batten (1992).

Comparative - a student's performance is compared with other students' performance, directly or indirectly.

Non-comparative - an individual's performances are assessed without reference to the standards or progress expected of others.

An even more common classification of assessment strategies, is to begin with norm-referenced and criterion-referenced assessment and set other strategies as originating from either of these, e.g. Cunningham (1986), Gronlund (1985), Ebel (1972).

Norm-referenced - performance is described in terms of an individual's relative standing in some defined group (e.g. spelt more test words correctly than 85 percent of the age group)

Criterion-referenced - describes what individuals can do within a specific domain and without reference to the performance of others (e.g. recognises as correct 17 of 20 spelling errors in the 200 word passage 'In Search of a Standard')

Elsewhere, norm-referenced and standards-based assessment are seldom seen as representing two ends of a continuum, unless the discussion is on comparative assessment only.

Depending on one's view of a broad assessment category, a continuum of assessment as suggested by Withers and Batten (1992), will either clarify or further confuse the relative place of standards-based assessment, or as they refer to it, standards-referenced assessment.

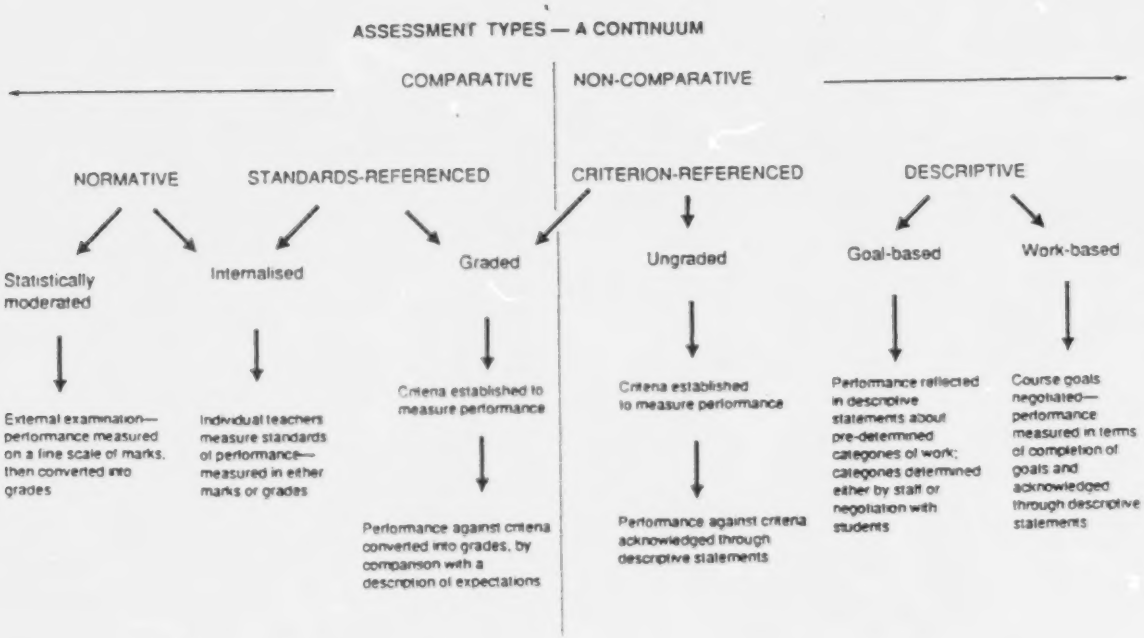
ASSESSMENT TYPES — A CONTINUUM
COMPARATIVE
NON-COMPARATIVE
NORMATIVE
STANDARDS-REFERENCED
CRITERION-REFERENCED
DESCRIPTIVE
Statistically moderated
Internalised
Graded
Ungraded
Goal-based
Work-based
External examination—performance measured on a fine scale of marks, then converted into grades
Individual teachers measure standards of performance—measured in either marks or grades
Criteria established to measure performance
Performance against criteria converted into grades, by comparison with a description of expectations
Criteria established to measure performance
Performance against criteria acknowledged through descriptive statements
Performance reflected in descriptive statements about pre-determined categories of work; categories determined either by staff or negotiation with students
Course goals negotiated—performance measured in terms of completion of goals and acknowledged through descriptive statements

Figure 2.1 Summary Table of the Comparable/Non-comparable Assessment Continuum

According to Withers and Batten (1992), this mode is set within the comparative sector. They go on to say,

> "In so far as the standards referred to are externally set, generally expected, and expressed as clear-cut grades, it shares some of the features of normative assessment. In so far as the standards expected are expressed as criteria, and student performance expressed in grade-related descriptors, it is synonymous with graded criterion-referenced assessment."

It is worth noting that all assessments from within the comparative sector of the continuum, have the potential to be a basis for competition between students for grades or awards. Furthermore we should consider whether standards-referenced assessment as outlined by Withers and Batten (1992) is similar in concept to standards-based assessment as conceived by NZQA.

Peddie (1992) p. 23 states,

> The term standards-based assessment is used when the measurement or outcome is assessed, in other words, "analysed", against some fixed criterion or level of achievement known as a "standard". A whole set of standards may be involved. These standards should be set in advance, so that they are well-known to both teachers and learners. In theory, each learner gets exactly what they achieve, so that it is possible - again in theory - for all learners to achieve the particular standard desired."

He notes that the number of learners who meet the standard will be determined quite substantially (and perhaps quite arbitrarily) by the level at which the standard is set. It is pointed out too, that in some instances features of the task itself will determine the standard i.e. safety considerations in a building course, acceptable tolerances in an engineering course, load and stress limitations in a building science course; but in other instances the standard will be based on, some expectations of what is achievable and thus some form of comparison. These expectations are to reflect:

1. experience of teachers and other experts
2. careful analysis of the unit and its learning outcomes.

Peddie (1992) goes on to indicate (p. 24) 'Neither the standards nor the final reported results depends on what a particular group taking the test happens to achieve'. This is satisfactory, as the point being made is that the proportion of candidates who

pass, for example, does not determine the value of the actual 'passing score'. However, it is quite wrong to assume that the setting of some standard which is not directly influenced by the performance of particular candidates, is not shaped, perhaps subtlely, by knowledge and expectations that teachers and other experts have developed about the performance of other learners. The point at issue is that in education, many examples of standards are based on expectations of what a reasonable performance level might be. Standards for many areas of education are more rooted in norm-based considerations than in considerations of the task itself.

In the earlier quote about standards-based assessment, (p. 2), competency-based and achievement-based assessment were given as major two sub-types.

Competency-based assessment is described in Peddie (1992) p, 24 as,

> "Where we set a particular standard which candidates must reach if they are to be judged as "competent", and therefore receive credit for the unit of learning...."
>
> "The standard here then, is a criterion level in specified skills or areas of knowledge. This is why competency-based assessment is also sometimes known as criterion-referenced assessment."

The fact that earlier NZQA publications did not link competency-based assessment with criterion-referenced assessment and standards-based assessment, did not advance clarity very much, particularly when Withers and Batten (1992) place standards-referenced assessment in the comparative side of their continuum and criterion-referenced assessment in the non-comparative side. Granted, graded standards-based assessment and graded criterion-referenced are seen as having common elements, albeit in the comparative zone.

Achievement-based assessment has been described also as another category of standards-based assessment. It is defined

by Peddie (1992), p. 26 as,

> "Assessment in which a number of progressively more demanding standards are used; and in which all learner achievement is reported, usually in the form of a number or letter grade."

> "Achievement-based assessment... is probably the type of standards-based assessment that teachers in secondary schools know best. ... sixth form trials used grade-related criteria as a way of arriving at an achievement-based assessment...."

A clear distinction is made with competency-based assessment, as here each learner will either meet or not meet some standard. Competency is demonstrated or it is not, thus competency is in part, determined by the standard which may be set in relation to some feature of the task, some comparative criteria related to examiner's knowledge, the level of performance that may be reasonably expected, or some marrying together of the two. What remains clear, is that determination of standards in some shape or form is at the heart of the matter.

Achievement-based assessment does not rest on a set of skills being present or absent, but on the degree to which they are present. Grade-related criteria, or graded standards on a five-point scale are central. In addition to the set of standards we have also to be concerned with the nature of the 5-point scale itself, including how results may be treated legitimately, when there is a need to combine or aggregate grades. We should ask too whether the grades represent an ordinal or interval scale? This is an important subsidiary consideration.

To return to the question posed earlier, what then is standards-based assessment within the NZQA model. On balance, standards-based assessment is portrayed with these characteristics:

- a form of assessment that contrasts with norm-referenced assessment

- a major category of assessment

- an assessment which is carried out against defined standards not the performance of others

- as a consequence of 'standard-referencing' there is no bar in theory, to every candidate achieving the standard or beyond

- competency-based assessment and achievement-based assessment are seen as major variants of standards-based assessment. Competency-based assessment is linked with criterion-referenced assessment, an almost universally-recognized term, but the link to achievement-based assessment is less clear

- the standard, a set of competencies, or a set of grade-related criteria are placed at the heart of standards-based assessment

- the standard, or the grade-related criteria, will not be influenced directly by the performance of examinees

- standards will reflect primarily, characteristics of the Unit of Learning, although it is noted as well, that expectations about performance may also be influential in determining standards.

The above sketch is put forward as a summary of the NZQA position. What is <u>not</u> acknowledged by this position is that:

- Norm-based assessment and standards-based assessment are portrayed as representing a dichotomy of assessment approaches. For most other commentators,

norm-based and criterion-based are the two broader categories.

- One type of standards-based assessment i.e. competency-based, has been linked with the cluster of techniques referred to as criterion-referenced assessment but the other type, i.e. achievement-based assessment has been left hanging.

- Standards-based assessment remains a comparative form of assessment in the view of some writers.

- Virtues claimed for standards-based assessment have been by negative reference to norm-based assessment, but without clear acknowledgement of the impact of comparative data in determining most educational standards.

- Little research has been undertaken on the characteristics or implementation of standards-based assessment in New Zealand This point is taken up again later.

- Little expertise in this area of assessment is likely to exist in our institutions.

The general notion of assessing to a standard makes eminent sense, and I for one would give total support to the principle. However, the issues are not quite as clear cut as they are being portrayed. In my view, it appears as though NZQA through much of their written material at least, have over-emphasised the criterion-referenced advantages of standards-based assessment and over-played the negative features of norm-referenced assessment, as a way of indicating why standards-based assessment is the desirable way to proceed. But, at the same time they have attempted to ignore the comparative nature of standards-based assessment within an education context and to down-play the

impact of normative information on the derivation of standards. Sass and Wagner (1992) in their discussion of norm-referencing and standards p. 20, suggest that norm-referencing is "... fundamentally unsuited for measuring student achievement in ways that are consistent with the Authority's Framework of Qualifications..."

My conclusion is that the requirements of the Framework have been the dominant force driving promotion of standards-based assessment, when questions of validity should have been to the forefront. No single assessment strategy is likely to provide every answer. Choice of an appropriate strategy for a particular context seems preferable.

The Question of Standards

Standards in some shape or form, presented as either criteria, descriptors, specified conditions, grade-related criteria, profiles of skills, achievement criteria, Peddie (1992); performance criteria, statements of performance, performance levels, criteria to determine achievement or competence, grading standard, Sass and Wagner (1992), are obviously central to standards-based assessment.

As a general notion in education, the concept of standards remains vexing. Croft (1991) and Croft (1992) has discussed this problem in relation to national monitoring studies, but there is some relevance to standards-based assessment as well.

In everyday speech we are comfortable with the term 'standard', when we state that 'the carpenter's work was of a good standard', 'standards of behaviour in public have changed', 'the standard of New Zealand cricket is better than it might have been'. In the management of our schools too, it is commonplace to expect our pupils to improve their 'standards of presentation', 'standards of behaviour', 'standards of speech', and their 'standards', fullstop.

What does this term mean? A dictionary definition of 'standard' is: "the degree of <u>excellence</u> required for particular purposes; the measure of what is <u>adequate</u>; a socially or

practically desired level of performance". Here we see that 'standard' may embrace simultaneously excellence, adequacy, and desirability.

In the sense that standard is used in these examples, 'the standard' represents some pre-conception of how well something should be accomplished; the skill with which an action might be performed; or the attitudes and behaviour one might reasonably expect of individuals in given situations. In other words, a standard is what we as individuals see as desirable, reasonable or appropriate.

When we discuss what learners actually achieve, we are in the area of norms. An important point to note is that norms and standards differ and that for standards-based assessment this difference is crucial. In the context of learning and teaching, standards may be regarded as objectives to be attained or perhaps expectations of desirable levels of performance. But average levels of performance may become internalised over time and come to be regarded as what may be expected. So the actual performance becomes confused with the standard.

An Introduction to the Framework may have further clouded the issue, I believe, by linking standards in a virtually undefined manner with diverse features such as learners' accomplishments (p. 2), assessment criteria (p. 5), efficiency of teaching methods (p. 5), quality of resources (p. 5), agreed outcomes (p. 6), absolute levels of achievement (p. 6), quality (p. 8), consistency (p. 12) and defining standards-based assessment as "assessment which is measured against unit standards" (p. 13) and unit standards as "published learning outcome statements and assessment criteria" (p. 14). Interestingly too, 'standard' is used no fewer than 18 times in the booklet but is not included in a 23 item glossary, although it appears in the definition of 7 terms within the glossary. It is interesting speculation as to whether standards within the Framework refers to excellence, adequacy, or desirability.

The foregoing illustrates again continuing difficulties with the term 'standards' and suggests its uselessness in educational measurement as an undefined term.

Warwick Elley (personal communication), has made a positive move to improve the utility of the term by distinguishing between a desired standard (a level at which one aims - excellence) and an obtained standard (a norm). He suggests too, that given the near impossibility of defining or reaching agreement on desired standards in most academic areas, that the matter of obtained standards may become a more realistic focus and could be viewed as an empirical question.

Elley (1993) has stated,

> "it should be stressed that educators in many other countries have tried to develop clear stand-alone standards in general subjects at the upper secondary level, but none has succeeded. Neither has NZQA. Many vocational topics do lend themselves to this model (e.g. can type 50 wpm, can weigh seeds accurately) but general subjects do not (e.g. can write a good persuasive essay, understands the causes of WWII, can carry on a conversation in French)."

> "At the heart of the problem in my view, is the NZQA assumption that standards can be spelled out in knowledge-based subjects, as if there were neat ladders of achievement in each curriculum area. In some aspects of mathematics, it may be feasible to list the skills to be mastered. But in English, social studies, science, the skills to be mastered are less important than and cannot be separated from, the knowledge they are applied to."

This is a view that should not be dismissed lightly.

Sass and Wagner in their independent report to NZQA (1992) give a clear lead on the matter of standards when they indicate (p. 25), "With the development of a standards-based system, the performance criteria that are used for grading purposes become the standards." Taken in conjunction with a later statement (p. 28) "... the statements of performance for those criteria themselves must necessarily be based on the normal range of achievement expected of the student in the programme", it is

apparent that they see standards as the performance criteria for awarding grades, and that these standards are in fact norms of achievement more than anything else.

There is little doubt that performance in relation to a standard could be measured, but before this is possible to an acceptable degree of reliability, the standard must be expressed in clear, easily interpreted terms. This is the nub of the issue.

Peddie (1992) does note a number of useful examples of standards being defined sufficiently to enable them to become a basis of measurement, but does obscure the issue a little by indicating that, "... School Certificate examination had at their heart a standard which was broadly speaking, related to an examination in which the average learner would score around fifty percent." The general line of reasoning linking this standard to the curriculum and expectations about what children may be expected to learn seems quite legitimate, but to interpret standards in this broad light is not consistent I believe, with the underlying notion of standards-based assessment. If this form of assessment is to approach the reliability associated with norm-referenced assessment, standards will need to be defined much more tightly than this.

Sass and Wagner (1992) also, note that norm-referenced measurement has an important role in selection, varying from competitive admission, to predicting academic success based on general academic and verbal skills. A second use is for developing the performance criteria for achievement-based assessment. A third is for diagnosing and monitoring. Underlying these three uses is the maximum consistency, hence reliability, that may be obtained from good quality norm-referenced measurement. It is this key area of reliability that standards-based assessment has yet to demonstrate.

Research and Development to Improve Standards-Based Assessment

What research and development seems called for to maximise the potential of standards-based assessment.

Sass and Wagner (1992) outline three matters they see as essential to underpin the success of standards-based assessment:

1. establishing the "normal range of achievement" so that grading criteria may reflect realistic expectations [sic] standards

2. maintaining on-going reviews of grading criteria so that these represent at all times, the expectations of the programme

3. maintaining consistency of grading by reviewing criteria, not by imposing a quota or using some other form of 'pressure' on teachers or tutors to conform to a 'desired' distribution of grades.

Clearly, a major and on-going research programme based on monitoring current achievement would be necessary to establish the "normal range of achievement", as current curricula are not sufficiently precise to provide grading criteria. Teacher or tutor judgement might be considered as an alternative to empirical research, but as indicated in Wagemaker (1993), standard-setting needs to take account of issues of time, the diverse background and experience of judges, and how differences that remain at the conclusion of a standard-setting exercise will be dealt with. It is worth recording that the exercise reported by Wagemaker (1993), involved a panel of 19 judges working for a full day on reading literacy at just two class levels.

Likewise, it is clear there are on-going costs and personnel resources in relation to Sass and Wagner's second and third proposals. More important I believe, are questions of validity in relation to the second proposal (what evidence apart from judgement, would there be that the criteria did in fact represent programme expectations?) and questions of reliability for the third proposal (what evidence is there that in every case consistent grading is achieved?)

In their earlier discussions on changing examination policy, Elley and Livingstone (1972) set out what they saw as the conditions needed for successful change. Although their document was in the context of moving from external examination to internal assessment, their five broad conditions provide a research agenda for standards-based assessment:

(i) validity of pupil assessments
(ii) maintaining parity of standards between schools and subject areas
(iii) acceptance by teachers who make the assessments
(iv) acceptance by institutions
(v) acceptance by the community at large

Elley and Livingstone (1972) see parity of standards as a central issue, and following data on variations of pass-rates for University Entrance from 162 schools, and variation in achievement on PAT: Reading Comprehension from 42 schools note, "These variations constitute a stubborn and unpalatable fact of life." They also ask, "How can such differences in standards be taken into account without a system of national examination?" The substance of the first part of their question is as pertinent as previously, although the reference to national examinations may now win less support. Also, within the current debate on standards, this statement refers more to norms, if we take standards to mean some combination of excellence, adequacy or desirability.

Nonetheless, their overview plus the elements of reliability, validity and recent advances in scaling would still provide the framework for research that is essential, if standards-based assessment is to function consistently and be interpreted with confidence. If NZQA have not instituted a programme of research along these general lines, they are remiss.

Establishing the qualities of validity and reliability of standards-based assessment, and of researching the conditions that will enhance these qualities within a standards-based environment, would seem to be a priority. This has been taken

up also by Peddie (1992) and Sass and Wagner (1992), who review a series of eight New Zealand studies on the general theme of establishing criteria and moderating final grades. In addition they cite six or so personal communications with anecdotal evidence.

In their useful overview and summary p. 38, they indicate that methods for ensuring parity include:

(a) the clear statement of objectives and grade-related criteria

(b) the use of more objective grading techniques with exemplary materials

(c) the forming of consensus panels of teachers, subject experts and industry representatives

(d) in-service training in methods of assessment and accreditation for teachers and tutors

(e) the more widespread use of subject or practitioner professional groups to be responsible for the setting up, maintenance and "ownership" of the standards for assessment

(f) setting up of item banks of standard test items where a sampling of the items can be used to determine the average ability level of a population, and

(g) the setting of common assessment tasks in an array of school-based assessment tasks.

They note that these techniques have been tried to some extent and have been useful in certain circumstances. By implication then, aspects of these techniques have been tried out, but some of the techniques have not been useful in certain

circumstances. This I think, underlines the state of current New Zealand research findings on assessment criteria and moderation. There is reasonably good understanding of the likely general requirements for defining criteria, but there is a definite lack of reliable knowledge of how these general requirements will be transformed into proficient strategies and then implemented validly and reliably in specific circumstances. It is a start for example, to conclude that "the use of more objective grading techniques with exemplary materials" are a prerequisite for valid and reliable assessment. However, much more specificity must be achieved by way of research, before this principle will result in satisfactory standards-based assessment. Principles have been identified, but their successful application may be a long way off. Sass and Wagner's further discussion of 19 Moderation strategies also illustrate the conclusion above, as again, the discussion remains at a general level.

Likewise, Walker (1990), concluded that some biology teachers could assess practical work reliably after training, and that some chemistry and physics teachers could not assess practical work reliably without training. Would this generalise to all teachers of biology, chemistry and physics? To all practical tasks within these disciplines? To all training for these science teachers? These are the sorts of questions that need to be answered before standards-based assessment may be used with confidence.

Conclusions

Some of the conflict identified with standards-based assessment seems to have come about because it has been promoted quite strongly as a form of non-comparative assessment, although elements of comparative assessment appear to be present. Additionally, NZQA's departure from widely accepted terminology has not helped. Nor has the way in which one example of standards-based assessment, namely competency-based assessment, has been equated with criterion-referenced assessment, while other forms have not. A framework such as the one used by

Withers and Batten (1992) would have been more helpful.

If standards-based assessment is to operate validly and reliably as a key element of the NZQA Framework, research into its particular features, strengths, and weaknesses would seem to be a priority. More importantly however, a stronger and more robust understanding of just how valid standards will be determined, and then applied in a reliable manner, to diverse sets of Unit Standards, by a range of staff, is needed also. It may need to be acknowledged that standards-based assessment procedures are not the most valid for all circumstances covered by the Framework, and other procedures may need to be investigated as well.

References

Croft, Cedric (1991) 'A Test For National Assessment'. Paper at Thirteenth National Conference, New Zealand Association for Research in Education, Dunedin, December 1991.

Croft, Cedric (1992) 'Who Might Benefit From National Monitoring?' Plenary Address to Annual Seminar of Otago Council, New Zealand Reading Association, Dunedin College of Education, March 1992.

Cunningham, G.K. (1986) Educational and Psychological Measurement, New York: Macmillan Publishing Company.

Ebel, R.L. (1972) Essentials of Educational Testing, Englewood Cliffs: Prentice Hall.

Elley, W.B. (1993) 'The NZQA Agenda Unfolds', The Christchurch Press. In press.

Elley, W.B. and Livingstone, I.D. (1972) External Examinations and Internal Assessments, Wellington: New Zealand Council for Educational Research.

Gronlund, N.E. (1985) Measurement and Evaluation in Teaching, New York: Macmillan.

New Zealand Qualifications Authority, An Introduction to the Framework, Wellington: The Authority.

Wagemaker, H. (Ed.) (1993) Achievement in Reading Literacy, Wellington: Ministry of Education.

Peddie, R. (1992) Beyond The Norm? An Introduction to Standards-based Assessment, Wellington: New Zealand Qualifications Authority.

Sass, R.E. and Wagner, G.A. (1992) Moderation Strategies For Achieving Consistency With Unit Standards - A Report to the New Zealand Qualifications Authority, Wellington: New Zealand Council for Educational Research.

Withers, G. and Batten, M. (1990) 'Defining Types of Assessment'. In B. Law and G. Withers (Eds.) Developments in School and Public Assessment, Melbourne: Australian Council for Educational Research.